



**University of
Zurich**^{UZH}

**Zurich Open Repository and
Archive**

University of Zurich
University Library
Strickhofstrasse 39
CH-8057 Zurich
www.zora.uzh.ch

Year: 2012

Evaluation of alternative RNA labeling protocols for transcript profiling with Arabidopsis AGRONOMICS1 tiling arrays

Müller, Marlen ; Patrignani, Andrea ; Rehrauer, Hubert ; Gruissem, Wilhelm ; Hennig, Lars

Abstract: Microarrays are routine tools for transcript profiling, and genomic tiling arrays such as the Arabidopsis AGRONOMICS1 arrays have been found to be highly suitable for such experiments because changes in genome annotation can be easily integrated at the data analysis level. In a transcript profiling experiment, RNA labeling is a critical step, most often initiated by oligo-dT-primed reverse transcription. Although this has been found to be a robust and reliable method, very long transcripts or non-polyadenylated transcripts might be labeled inefficiently. In this study, we first provide data handling methods to analyze AGRONOMICS1 tiling microarrays based on the TAIR10 genome annotation. Second, we describe methods to easily quantify antisense transcripts on such tiling arrays. Third, we test a random-primed RNA labeling method, and find that on AGRONOMICS1 arrays this method has similar general performance as the conventional oligo-dT-primed method. In contrast to the latter, however, the former works considerably better for long transcripts and for non-polyadenylated transcripts such as found in mitochondria and plastids. We propose that researchers interested in organelle function use the random-primed method to unleash the full potential of genomic tiling arrays.

DOI: <https://doi.org/10.1186/1746-4811-8-18>

Posted at the Zurich Open Repository and Archive, University of Zurich

ZORA URL: <https://doi.org/10.5167/uzh-71723>

Journal Article

Published Version



The following work is licensed under a Creative Commons: Attribution 2.0 Generic (CC BY 2.0) License.

Originally published at:

Müller, Marlen; Patrignani, Andrea; Rehrauer, Hubert; Gruissem, Wilhelm; Hennig, Lars (2012). Evaluation of alternative RNA labeling protocols for transcript profiling with Arabidopsis AGRONOMICS1 tiling arrays. *Plant Methods*, 8:18.

DOI: <https://doi.org/10.1186/1746-4811-8-18>



METHODOLOGY

Open Access

Evaluation of alternative RNA labeling protocols for transcript profiling with Arabidopsis AGRONOMICS1 tiling arrays

Marlen Müller¹, Andrea Patrignani², Hubert Rehrer², Wilhelm Gruissem^{1,2} and Lars Hennig^{1,3*}

Abstract

Microarrays are routine tools for transcript profiling, and genomic tiling arrays such as the Arabidopsis AGRONOMICS1 arrays have been found to be highly suitable for such experiments because changes in genome annotation can be easily integrated at the data analysis level. In a transcript profiling experiment, RNA labeling is a critical step, most often initiated by oligo-dT-primed reverse transcription. Although this has been found to be a robust and reliable method, very long transcripts or non-polyadenylated transcripts might be labeled inefficiently. In this study, we first provide data handling methods to analyze AGRONOMICS1 tiling microarrays based on the TAIR10 genome annotation. Second, we describe methods to easily quantify antisense transcripts on such tiling arrays. Third, we test a random-primed RNA labeling method, and find that on AGRONOMICS1 arrays this method has similar general performance as the conventional oligo-dT-primed method. In contrast to the latter, however, the former works considerably better for long transcripts and for non-polyadenylated transcripts such as found in mitochondria and plastids. We propose that researchers interested in organelle function use the random-primed method to unleash the full potential of genomic tiling arrays.

Keywords: Transcript profiling, Arabidopsis, Microarray, AGRONOMICS1, Tiling array, RNA labeling

Background

Transcript profiling has become a routine experimental approach in many fields of biology. One way to perform such profiling is to use DNA microarrays. For the model plant *Arabidopsis thaliana*, microarrays that probe the transcriptome have been used for more than ten years [1-8], and ATH1 arrays manufactured by Affymetrix have probably been used most widely [9]. Although ATH1 arrays proved to generate robust and reliable data, they lack probes for one third of the annotated Arabidopsis genes. For genome-wide profiling, researchers have started to use transcriptome arrays from other manufacturers or genome tiling arrays, which contain probes against the entire genome. Tiling arrays are not restricted to probe mRNA but also other transcripts such as sRNA (small RNA), tRNA (transfer RNA) and

miRNA (micro RNA), and yield information on splicing. In Arabidopsis, tiling arrays have already identified many novel genes, intergenic non-coding RNAs and antisense transcripts [10-16]. Tiling arrays have also frequently been used in Arabidopsis epigenome profiling such as with chromatin-immunoprecipitation (ChIP-chip) (for review see [17]) and for detection of deletion mutations [18]. Another major advantage of tiling arrays is that they are not limited to current genome annotations but can be re-analyzed when new genome annotation information becomes available. In *Arabidopsis thaliana*, the Affymetrix Arabidopsis 1.0R tiling array was used in diverse applications ranging from transcript discovery to ChIP-chip [13,15,19-23]. An alternative Arabidopsis tiling array is the recently developed AGRONOMICS1 Affymetrix array [24]. The design of the AGRONOMICS1 array is similar to the Affymetrix Arabidopsis 1.0R array but lacks mismatch probes. Instead, it contains probes against both genome strands while the Affymetrix Arabidopsis 1.0R array probes only one strand. This allows strand-specific transcriptome profiling on

* Correspondence: lars.hennig@slu.se

¹Department of Biology, Plant Biotechnology, ETH Zurich, Zurich, Switzerland

³Department of Plant Biology and Forest Genetics, Uppsala BioCenter, Swedish University of Agricultural Sciences and Linnean Center for Plant Biology, PO-Box 7080SE-75007, Uppsala, Sweden

Full list of author information is available at the end of the article

single AGRONOMICS1 arrays and gives doubled probe density for epigenome profiling applications.

Since its release, AGRONOMICS1 arrays have been used for many experiments some of which have been published [25-27]. Initially, the *GeneChip®3' IVT Express kit* (Affymetrix) was utilized to prepare samples for hybridization to AGRONOMICS1 arrays because this kit was also widely used to prepare samples for ATH1 arrays. Nevertheless, this method has some disadvantages, mostly because it is based on oligo-dT priming. Oligo-dT priming disfavors labeling of long transcripts. In addition, non-polyadenylated transcripts are labeled only poorly. Because transcripts in organelles usually lack polyA tails [28], oligo-dT-based labeling methods are not suitable for projects where expression data for plastidial or mitochondrial genes are needed. In addition, it has been reported that the use of T7 sequences in common oligo-dT priming protocols can cause artifacts on tiling arrays [29]. Labeling methods based on random priming at the reverse transcription step carry the potential to overcome the limitations of oligo-dT-based labeling [30-34].

Here, we tested an alternative labeling method based on random priming at the reverse transcription step and develop appropriate data analysis routines. The comparison of this and the previously used protocols revealed, in general, a very good agreement between fold change values from both methods. Considerable differences between both methods were observed for transcript signal estimates for long transcripts and for organellar transcripts, which both are only inefficiently labeled by oligo-dT priming. In both cases, expression estimates were much larger for the method based on random priming. In summary, we present an RNA labeling procedure together with an appropriate data analysis pipeline that can replace established oligo-dT based methods in projects where expression of plastidial or mitochondrial genes is of interest.

Results and discussion

Analyzing AGRONOMICS1 arrays based on the TAIR10 genome version

A major advantage of genome tiling microarrays is that they can accommodate changes in genome annotation.

After the AGRONOMICS1 array was developed, the TAIR10 version of the Arabidopsis genome was released. We generated new TAIR10-based CDF (chip description format) files, which can be used to generate expression estimates from raw data (CEL files) not only for new but also for past experiments (Table 1). We compared expression estimates derived from a published data set from dark-grown and illuminated seedlings, which was based on labeling with the Affymetrix IVT Express kit [24], using the TAIR9- and the TAIR10-based CDF files, and as expected found very high agreement between both (Additional file 1: Figure S1). Because antisense transcripts have attracted much attention in recent years, we generated a second TAIR10-based CDF file that can be used to profile antisense transcripts. Note that originally most labeling protocols for Affymetrix expression arrays generated labeled aRNA (antisense RNA), which hybridizes with probes against the antisense strand. Nowadays, however, protocols generating labeled cDNA or labeled aRNA are both common. Table 2 summarizes usage of the new CDF files.

Applying the new CDF files to the dataset of RNA from dark-grown and illuminated seedlings, we identified 780 and 333 genes that were induced or repressed by light, respectively ($p < 0.01$, fold change > 2). With the same criteria, 5 and 0 genes had no significant difference in abundance of sense transcripts but had antisense transcripts that were induced or repressed by light, respectively. Three of the 5 genes overlapped with annotated genes on the opposite genome strand. Visual inspection of the tiling array data suggested that in these cases the apparent antisense signal was caused by a sense signal from the overlapping gene. In contrast, for two genes (*AT4G31875* and *AT5G64401*) strong antisense signals could not be explained by an overlap with known genes (Figure 1). Note that the IVT Express labeling protocol used here was earlier shown to have high strand-specificity [24]. Thus, the new antisense-CDF file can be used to quantify antisense transcripts.

The approach presented here differs from that of Coram and colleagues, who also quantified antisense transcript [35]. These authors used Affymetrix GeneChip

Table 1 Properties of custom-made CDF files for AGRONOMICS1 arrays

CDF file	Genome version	# of probes (ath ¹⁾ + control probes)	# of probe sets (ath + control probe sets)	# of genes probed(M + P) ²⁾	Description
<i>agronomics1_ TAIR9_gene</i>	TAIR9(2009)	1,246,484 (1,093,816 + 152,668)	30,608 (30,237 + 371)	29,920(33 + 67)	Gene-specific probe sets of all suitable probes from the antisense strand
<i>agronomics1_ TAIR10_gene</i>	TAIR10(2010)	1,626,780 (1,474,112 + 152,668)	31,027 (30,656 + 371)	30,466(35 + 85)	Gene-specific probe sets of all suitable probes from the antisense strand
<i>agronomics1_ TAIR10_gene_sense</i>	TAIR10(2010)	1,397,929 (1,245,261 + 152,668)	30,757 (30,386 + 371)	30,174(36 + 84)	Gene-specific probe sets of all suitable probes from the sense strand

¹⁾ ath probes, probes against Arabidopsis sequences.

²⁾ M, mitochondrial genes; P, plastidial genes.

Table 2 Matching CDF files and RNA labeling protocols

CDF file	3'-IVT kit	WT kit
<i>agronomics1_TAIR10_gene</i>	Detection of sense transcripts	Detection of antisense transcripts
<i>agronomics1_TAIR10_gene_antisense</i>	Detection of antisense transcripts	Detection of sense transcripts

Wheat Genome arrays, which are 3'IVT expression arrays and carry probes only for the sense strand. Therefore, two alternative labeling methods were used to label transcripts derived from sense or antisense transcription, respectively. Labeled samples were separately hybridized to arrays making two arrays per sample needed. It also required sufficient RNA for two labeling reactions per sample, which could be limiting for rare samples. In addition, the different labeling protocols imposed different sensitivities. In contrast, our approach relies on only one labeling reaction and hybridization and does not increase required amounts of RNA or experimental costs. Instead, labeled antisense transcripts are directly probed by complementary oligonucleotides present on the AGRONOMICS1 array. Because different probes are used to interrogate sense and antisense transcripts, signal intensities can also in this case not directly be compared. In most cases,

however, such probe-specific effects will have a minor impact on the expression signals generated by RMA. It should also be noted that this approach approximates potential antisense transcripts based on the annotation of the sense transcript. To accurately determine transcript boundaries, algorithms to segment the hybridization signal along chromosomal coordinates are needed [36].

Performance of the oligo-dT-based and random-primed labeling protocols

Aliquots from the same RNA preparations that were used previously with an oligo-dT-based protocol [24] were used for the random-primed protocol. Two technical replicates of a rosette leaf RNA sample and three technical replicates of a flower RNA sample were labeled and hybridized to AGRONOMICS1 tiling arrays. Table 3 shows the correlation among the technical replicates from the oligo-dT-based and the random-primed protocols. Although both protocols resulted in high data concordance, the random-primed protocol generated data with a slightly higher correlation.

Next, we tested how well results based on the two labeling protocols correlated with each other. As evident from Figure 2, correlation between replicates of the same protocol was considerable higher than correlation between replicates of different protocols. This result suggested that array-based expression signals differed for many

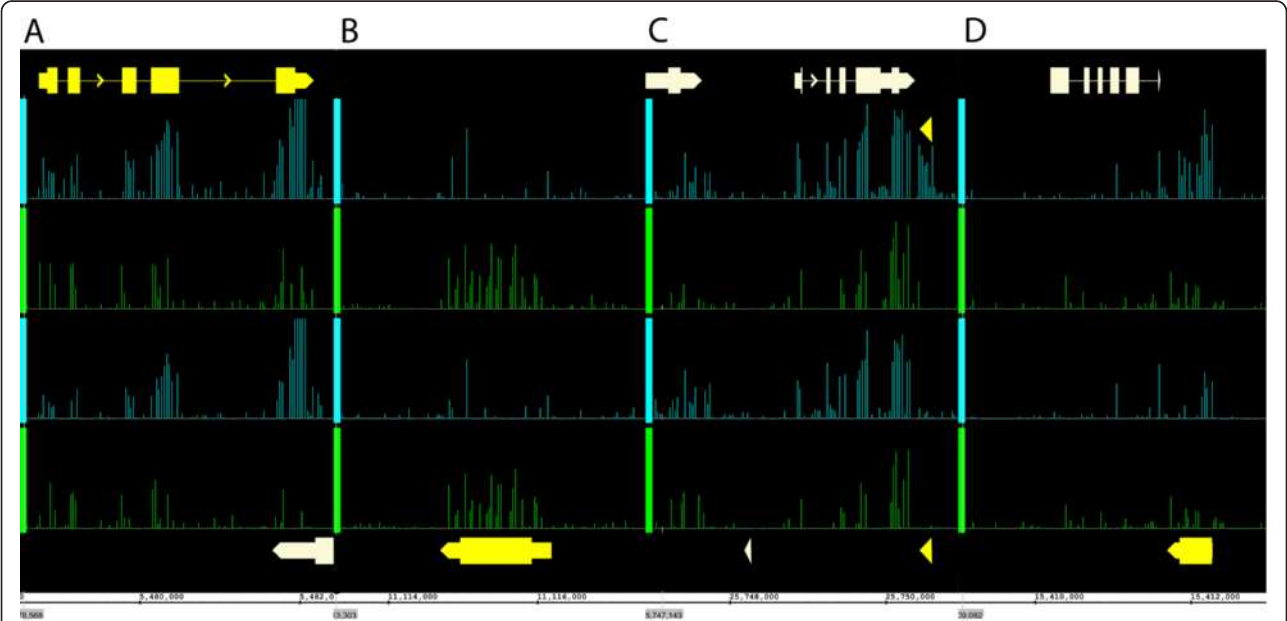


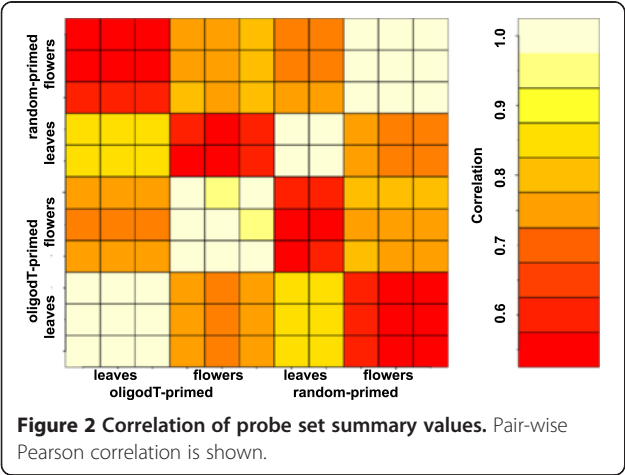
Figure 1 Detection of antisense transcript signals on AGRONOMICS1 arrays. Bars represent gene annotations. Genes of interest are yellow, other genes grey. Probe signals from the *plus* strand are cyan and from the *minus* strand green. The upper two rows of signals show RNA from irradiated seedlings, the lower two rows RNA from dark-grown seedlings. **(A)** *AT1G15950*, a gene on the *plus* strand. Signals are highest for probes from annotated exons and on the *plus* strand. **(B)** *AT1G31130*, a gene on the *minus* strand. Signals are highest for probes on the *minus* strand. **(C-D)** *AT5G64401* **(C)** and *AT4G31875* **(D)** are genes on the *minus* strand. However, signals are highest on the *plus* strand strongly suggesting antisense transcription. Note that the signals are specific for irradiated seedlings.

Table 3 Correlation of probe set summary signals

	oligo-dT primed	random primed
<i>Index.R</i>		
Leaves	0.815 ± 0.023	0.861 ± 0.009
Flowers	0.889 ± 0.011	0.920
<i>C_{Pearson}</i>		
Leaves	0.982 ± 0.003	0.977 ± 0.004
Flowers	0.962 ± 0.004	0.991

Mean and S.D. are listed for technical replicates. Correlation was expressed as *index.R* value [37] and as pair-wise Pearson correlation coefficient.

genes. Because transcripts in plastids and mitochondria are only polyadenylated as part of a polyadenylation-dependent RNA degradation mechanism [28], we hypothesized that expression signals for organellar genes would differ most between the two labeling protocols. Consistent with this hypothesis, there were mostly small differences between expression signals from nuclear genes while plastidial and in particular mitochondrial genes had consistently much higher expression signals when using the random-primed labeling protocol (Figure 3A). Expression signals of mitochondrial genes were independently of the protocol similar between flowers and leaves but about sixteen times larger when using the random-primed labeling protocol (Figure 3B). While signals were usually close to the detection limit when using the oligo-dT-primed method, many mitochondrial transcripts gave signals that were among the strongest in the genome when using the random-primed method. Expression signals of plastidial genes were independently of the protocol higher in leaves than in flowers. Nevertheless, these signals were about eight times larger when using the random-based labeling protocol and among the highest signals obtained on this array (Figure 3C). Signals for nuclear transcripts, in contrast, did not globally differ strongly between flowers and leaves, and also the labeling protocol had only a mild effect on these genes (Figure 3D).



In addition to organellar transcripts, also very long transcripts are expected to yield signals that differ particularly strongly between the labeling protocols, because multiple priming events in a random-primed protocol will generate more cDNA than single priming events in an oligo-dT-primed protocol. We tested this hypothesis by grouping the nuclear genes in 20 bins according to transcript lengths and plotting signal differences between protocols separately for each bin for flowers (Figure 4A) and leaves (Figure 4B). Indeed, signals for the ~25% shortest transcripts proved to be independent of the labeling protocol while signals for longer transcript were usually considerably larger when using the random-primed protocol. In contrast, differences in signal strength between leaves and flowers were independent of transcript length regardless of the labeling protocol (Figure 4C,D). These results show that expression signals are strongly affected by the labeling protocol, and thus direct comparisons of expression signals from experiments using different labeling protocols should be avoided. In contrast to expression signals, signal ratios did not strongly depend on transcript length (Additional file 1: Figure S2) suggesting that signal ratios can be compared even between experiments using different labeling methods. Because long transcripts can be interrogated by more probes than short transcripts, tiling array-based expression signals for long transcripts are expected to have higher precision than the signals for short transcripts. This effect is clearly visible in data based on oligo-dT priming for the shortest transcripts (Figure 5A,B). In contrast, for transcripts of intermediate length no such effect is evident, and for the longest transcripts signal variability even increases considerably. This increase in signal variability indicates variable labeling efficiency for long transcripts. In contrast, variability of expression signals based on random priming decreases over almost the entire range of transcript sizes (Figure 5C,D), suggesting that a major effect of reduced labeling efficiency of oligo-dT priming is increased measurement variability.

Finally, we tested whether the used random-primed labeling method has sufficient strand-specificity to allow simultaneous detection of sense and antisense transcripts as shown above for data based on oligo-dT-primed labeling. Plotting sense and antisense signals for each gene revealed that even for genes with high sense transcript signals the antisense signal was usually very low (Figure 6A). For comparison, a plot of sense signals from leaf and flower samples showed a high correlation despite the great difference in tissue composition (Figure 6B). These results show that the used random-primed labeling method has a sufficient strand-specificity to justify quantification of antisense transcription.

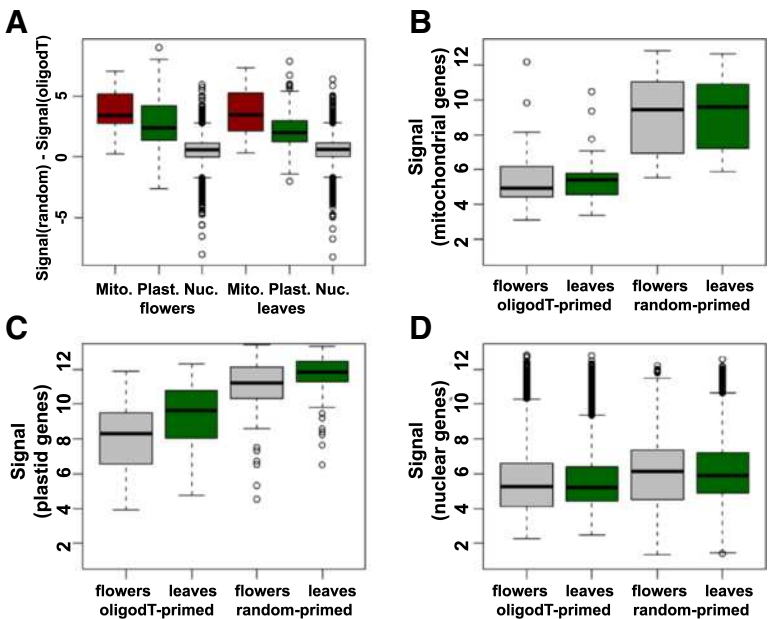


Figure 3 Expression signals of organellar genes. (A) Differences of signals from oligo-dT and random priming. Values larger than 0 indicate higher signals when using random priming. Mitochondrial, plastidial and nuclear genes were analyzed separately. Note that RMA-derived expression signals are in logarithmic scale. Therefore, differences in RMA-derived signals are equivalent to log2 of ratios (i.e. fold change) of signals in linear scale. (B-D) Expression signals in flowers (grey) and leaves (green) for mitochondrial (B), plastidial (C) and nuclear genes (D).

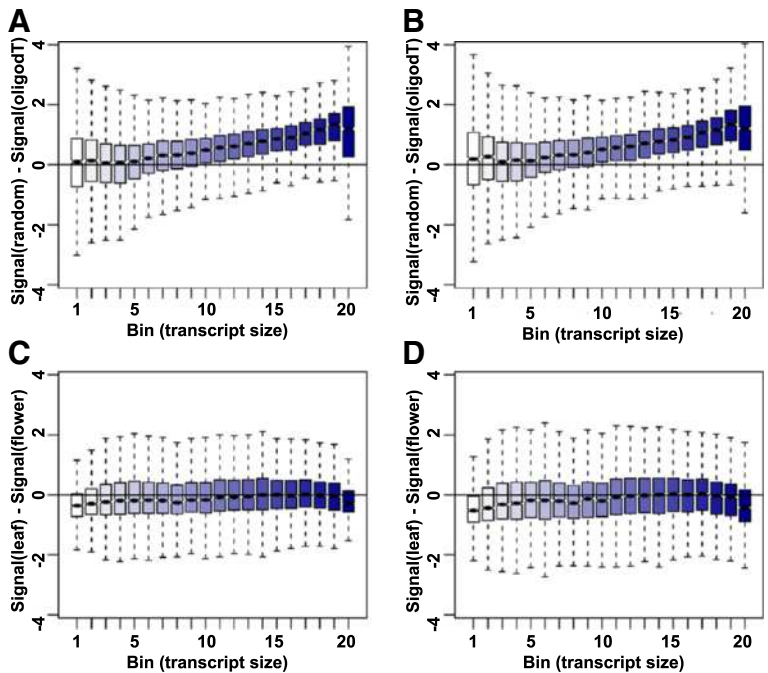


Figure 4 Effect of gene length on expression signals. (A-B) Differences of signals based on oligo-dT and random priming. Values larger than 0 indicate higher signals when using random priming. Nuclear genes were sorted into 20 bins according to transcript length. Signals from flower RNA (A) and leaf RNA (B). (C-D) Differences of signals from flowers and leaves. Values larger than 0 indicate higher signals in leaf RNA. Signals are based on oligo-dT (C) and on random priming (D).

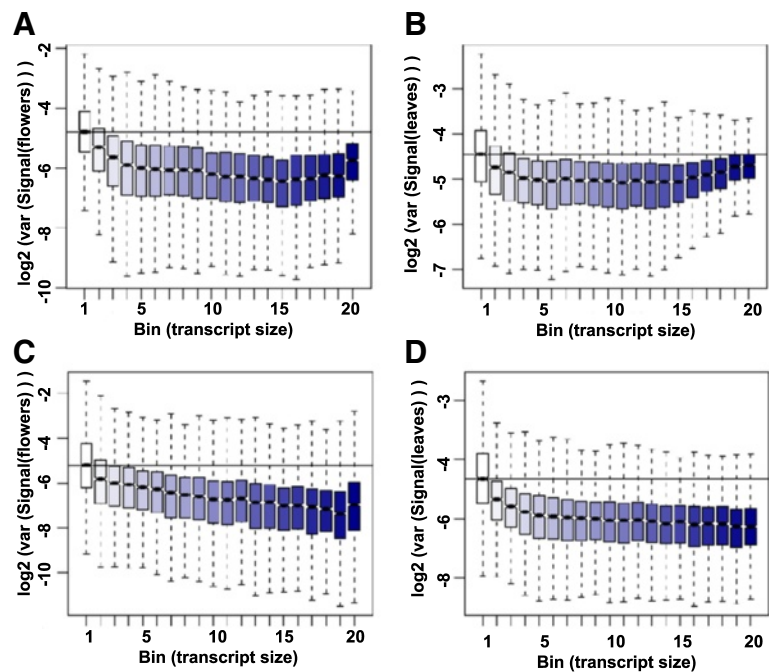


Figure 5 Effect of gene length on expression variance. Log₂ of signal variance for oligo-dT (A-B) and random priming (C-D). Nuclear genes were sorted into 20 bins according to transcript length.

Together, we found that the random-primed labeling protocol performed similar to the oligo-dT-primed protocol in most comparisons, but was more sensitive for organellar and long transcripts.

Conclusions

Genomic tiling microarrays are valuable tools for biology, and we have developed two extensions that expand the application range of AGRONOMICS1 tiling arrays.

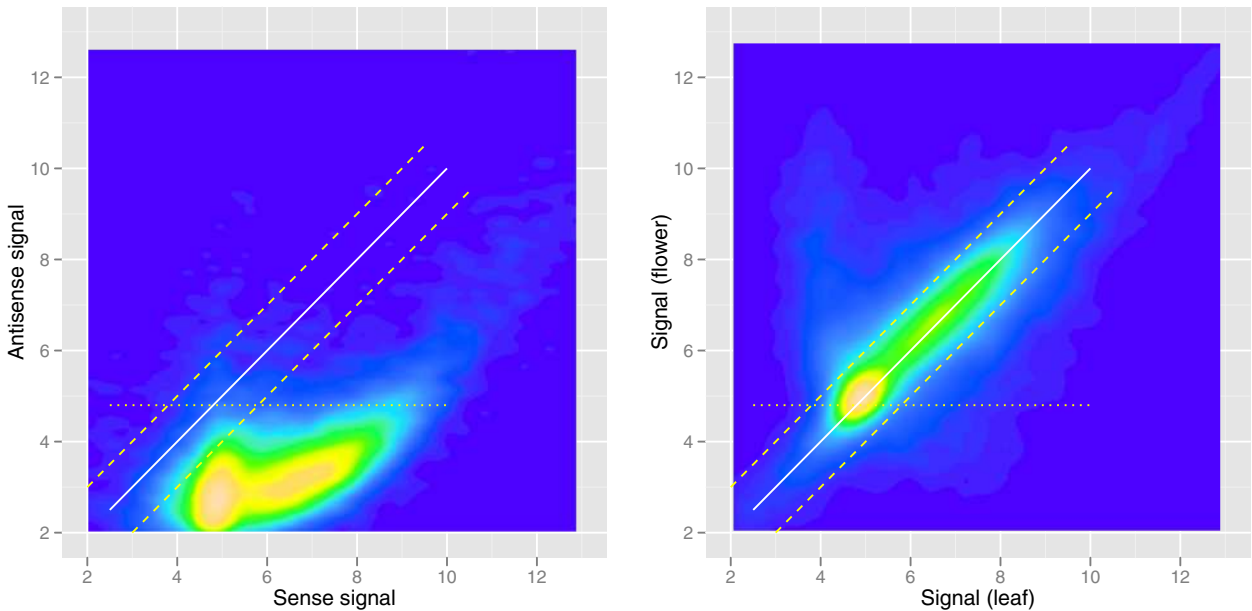


Figure 6 Strand specificity of sample labeling. (A) Leaf RNA was labeled using random priming and hybridized to AGRONOMICS1 arrays. Sense and antisense expression signals were calculated from the same CEL files using alternative CDF files. (B) Leaf and flower RNA was labeled using random priming and hybridized to AGRONOMICS1 arrays, and sense expression signals were calculated. Lines in (A) and (B) indicate: signal identity ($x=y$), solid lines; fold change of +2 and -2, dashed lines; empirical threshold for background signals, dotted lines.

First, we developed new CDF files for these arrays. Because the CDF files are based on the latest Arabidopsis genome version (TAIR10), estimation of gene expression levels will be more reliable. Importantly, raw data from past experiments can easily be re-analyzed with the new files. The AGRONOMICS1 tiling array contains probes from both genome strands, and we developed CDF files that contain probes located within annotated exons and match either the sense or the antisense strand. The CDF files can be used to simultaneously estimate levels of sense and antisense transcripts without the need for additional experiments or array hybridization.

Second, we tested an alternative labeling protocol that is not based on oligo-dT priming. Oligo-dT-based labeling methods are reliable and widely used for transcript profiling, but they suffer from certain deficiencies. In particular, oligo-dT priming fails to efficiently label organellar and very long transcripts. We found that when using AGRONOMICS1 tiling arrays a random-primed protocol compares favorably to the conventional oligo-dT-primed protocol. First, reproducibility of technical replicates was similar or even higher for the random-primed protocol. In addition, signal log ratios did not globally differ between both labeling methods, indicating that overall results are consistent and comparable. In contrast to signal ratios, signal values were less similar between the two methods. Therefore, a direct comparison of expression values is only justified for one and the same labeling method. Second, expression signals for long transcripts were considerably higher when using random priming. This causes an improved signal-to-noise ratio specifically for long transcripts. Third, expression signals of organellar transcripts were detected with much greater sensitivity and greater precision when using random priming.

In summary, alternative CDF files or labeling protocols enable the utilization of AGRONOMICS1 tiling arrays to interrogate antisense transcripts, transcripts from organelles or transcripts of very long genes in addition to the commonly probed nuclear mRNAs.

Methods

Plant material and RNA extraction

RNA samples were as described [24]. Briefly, *Arabidopsis thaliana* accession Columbia-0 plants were grown on soil at 23 °C in a photoperiod of 16 h of light and 8 h of darkness. Leaves (no. 4 from 10–15 plants per sample) and flowers (stage 15; 20–25 per sample) were harvested after 10 and 25 d, respectively. Total RNA was isolated using the Qiagen Plant RNeasy MiniKit according to the manufacturer's instructions.

Microarray target preparation

Method 1. GeneChip® IVT express kit

Microarray target preparation with the GeneChip® IVT Express Kit (Affymetrix, Santa Clara, CA) was described before [24].

Method 2. GeneChip® whole transcript (WT) sense target labeling assay

The starting material was 1 µg of total RNA. Then, microarray target preparation with the GeneChip® Whole Transcript (WT) Sense Target Labeling Assay (Affymetrix, Santa Clara, CA) was carried out as recommended by the manufacturer. Briefly, ribosomal RNA was removed using a RiboMinus™ Plant Kit (Invitrogen, Zug, Switzerland), which is not dependent on the polyadenylation status or the presence of 5'cap structure on the RNA. Then, random hexamers tagged with T7 promoter sequence are used to conduct a two-cycle cDNA synthesis following.

Array hybridization

Biotin-labeled microarray target samples were mixed in 300 µl of Hybridization Mix (Affymetrix) containing Hybridization Controls and Control Oligonucleotide B2 (Affymetrix). Samples were hybridized onto Affymetrix AGRONOMICS1 Arabidopsis tiling array for 16 h at 45 °C. Arrays were then washed using an Affymetrix Fluidics Station 450 following the FS450_0004 protocol. An Affymetrix GeneChip Scanner 3000 was used to measure the fluorescence intensity emitted by the labeled target.

Generation of CDF files

Custom-made CDF files were generated as described [24]. Briefly, probes were mapped to the TAIR 10 genome sequence, and only probes with a single match inside an annotated exon (excluding untranslated regions) were used. Probe sets were generated if at least three such probes existed for a gene. For genes with multiple transcripts with little overlap, more than one probe set was generated per gene (see Table 4). The CDF file contains three types of probe sets, which can be discriminated by their names. The naming scheme is <locus name>.<variant>.<chromosome>.<strand>.<mRNA_start>.<mRNA_end> (e.g. AT1G01010.0.Chr1.plus.3631.5899). The meaning of the variant component is as follows: 0, there is only one transcript annotated for the gene, and the probe set matches this transcript; X, there are multiple transcripts with a large overlap annotated for the gene, and the probe set matches the intersection of all these transcripts; [1,N], there are multiple transcripts with little overlap annotated for the gene, and each probe set contains all probes that match the corresponding. Finally, there is a number

Table 4 Summary of probe sets on custom-made CDF files for AGRONOMICS1 arrays

	<i>agronomics1_</i> <i>TAIR9_gene</i>	<i>agronomics1_</i> <i>TAIR10_gene</i>	<i>agronomics1_</i> <i>TAIR10_gene_antisense</i>
Number of probe sets	30,608	31,027	30,757
Control probe sets	371	371	371
Probe sets for Arabidopsis genes	30,237	30,656	30,386
Probe sets for nuclear genes	30,137	30,536	30,266
Probe sets for plastid genes	67	85	84
Probe sets for mitochondrial genes	33	35	36
One annotated transcript uniquely probed	25,387	24,426	24,139
The intersection of a set of annotated transcripts uniquely probed	4,325	5,661	5,644
Several annotated transcript exist with little overlap; 1 is specifically probed	0	90	160
Several annotated transcript exist with little overlap; 2 are specifically probed	131	70	86
Several annotated transcript exist with little overlap; 3 are specifically probed	54	36	44
Several annotated transcript exist with little overlap; 4 are specifically probed	15	6	6
Several annotated transcript exist with little overlap; 5 are specifically probed	7	1	4
Several annotated transcript exist with little overlap; 6 are specifically probed	1	1	1
2 largely overlapping genes probed	0	90	85
3 largely overlapping genes probed	0	2	2
4 largely overlapping genes probed	0	2	2

of overlapping genes annotated in the genome for which no gene-specific probes sets could be formed. Thus, 94 probe sets were included that probe more than one annotated gene (90 probe 2 genes, 2 probe 3 genes, 2 probe 4 genes; e.g. AT1G06149_AT1G06150.X.Chr1.minus.1867015.1873718).

Data analysis

All analysis was performed in R 2.12.1 [38]. Visualization of tiling array data was done using the Integrated Genome Browser at <http://igb.bioviz.org> [39]. Library files and scripts are freely available as Supplemental Data, at http://www.agron-omics.eu/index.php/resource_center/tiling-array or upon request from the authors. Expression signals were extracted from CEL files using RMA [40] implemented in the *aroma.affymetrix* package [41] as described earlier [24]. For the comparison of labeling methods, only genes with unique probe sets in both CDF files were used. Quantile normalization as implemented in the *limma* package [42] was used for normalization of expression values to achieve consistency between arrays.

Availability

Library files and scripts are freely available as Supplemental Data S1 at www.agron-omics.eu/index.php/resource_center/tiling-array and at www.slu.se/genetics/resources/agronomics1

Additional file

Additional file 1: Figures S1. Correlation of expression signals from TAIR9- and TAIR10-based CDF files. Correlation is measured with *index.R*. The axis label *sig1* denotes replicates of TAIR9, *sig2* replicates of TAIR10. F1, F2, F3 are the flower replicates; L1, L2, L3 are the leaf replicates. **(A)** oligo-dT priming, **(B)** random priming. **Figure S2.** Signal ratios are not affected by gene length. Signal log ratios (SLR) between flowers and leaves were calculated, and differences between the SLR values based on oligo-dT and random priming were plotted. Values larger than 0 indicate higher fold changes when using oligo-dT priming. Nuclear genes were sorted into 20 bins according to transcript length.

Competing interests

The authors declare that they have no competing interests.

Acknowledgements

We thank Alessandro Fammartino and Remy Bruggmann (Functional Genomics Center Zurich) for help with array hybridizations and generating the CDF files, respectively, and Sascha Laubinger (University Tübingen) for providing RNA. This work was supported by the Sixth Framework Program of the European Commission through the AGRON-OMICS Integrated Project (grant no. LSHG-CT-2006-037704).

Author details

¹Department of Biology, Plant Biotechnology, ETH Zurich, Zurich, Switzerland. ²Functional Genomics Center Zurich, ETH and University of Zurich, CH-8057, Zurich, Switzerland. ³Department of Plant Biology and Forest Genetics, Uppsala BioCenter, Swedish University of Agricultural Sciences and Linnean Center for Plant Biology, PO-Box 7080SE-75007, Uppsala, Sweden.

Authors' contribution

LH and WG discussed and LH designed the experiments. AP performed the experiments. MM, HR and LH analyzed the data. MM, LH and WG wrote the manuscript. All authors read and approved the final manuscript.

Received: 31 January 2012 Accepted: 13 June 2012

Published: 13 June 2012

References

- Hilson P, Allemeersch J, Altmann T, Aubourg S, Avon A, Beynon J, Bhalerao RP, Bitton F, Caboche M, Cannoot B, *et al*: Versatile gene-specific sequence tags for Arabidopsis functional genomics: transcript profiling and reverse genetics applications. *Genome Res* 2004, **14**:2176–2189.
- Redman JC, Haas BJ, Tanimoto G, Town CD: Development and evaluation of an Arabidopsis whole genome Affymetrix probe array. *Plant J* 2004, **38**:545–561.
- Ma L, Chen C, Liu X, Jiao Y, Su N, Li L, Wang X, Cao M, Sun N, Zhang X, *et al*: A microarray analysis of the rice transcriptome and its comparison to Arabidopsis. *Genome Res* 2005, **15**:1274–1283.
- Schena M, Shalon D, Davis RW, Brown PO: Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* 1995, **270**:467–470.
- Reymond P, Weber H, Damond M, Farmer EE: Differential gene expression in response to mechanical wounding and insect feeding in Arabidopsis. *Plant Cell* 2000, **12**:707–720.
- Schenk PM, Kazan K, Wilson I, Anderson JP, Richmond T, Somerville SC, Manners JM: Coordinated plant defense responses in Arabidopsis revealed by microarray analysis. *Proc Natl Acad Sci U S A* 2000, **97**:11655–11660.
- Wisman E, Ohlrogge J: Arabidopsis microarray service facilities. *Plant Physiol* 2000, **124**:1468–1471.
- Hennig L, Menges M, Murray JAH, Gruissem W: Arabidopsis transcript profiling on Affymetrix genechip arrays. *Plant Mol Biol* 2003, **53**:457–465.
- Zimmermann P, Hirsch-Hoffmann M, Hennig L, Gruissem W: Genevestigator. Arabidopsis microarray database and analysis toolbox. *Plant Physiol* 2004, **136**:2621–2632.
- Zhang X, Borevitz JO: Global analysis of allele-specific expression in Arabidopsis thaliana. *Genetics* 2009, **182**:943–954.
- Stolc V, Samanta MP, Tongprasit W, Sethi H, Liang S, Nelson DC, Hegeman A, Nelson C, Rancour D, Bednarek S, *et al*: Identification of transcribed sequences in Arabidopsis thaliana by using high-resolution genome tiling arrays. *Proc Natl Acad Sci U S A* 2005, **102**:4453–4458.
- Laubinger S, Sachsenberg T, Zeller G, Busch W, Lohmann JU, Ratsch G, Weigel D: Dual roles of the nuclear cap-binding complex and SERRATE in pre-mRNA splicing and microRNA processing in Arabidopsis thaliana. *Proc Natl Acad Sci U S A* 2008, **105**:8795–8800.
- Laubinger S, Zeller G, Henz SR, Sachsenberg T, Widmer CK, Naouar N, Vuylsteke M, Scholkopf B, Ratsch G, Weigel D, *et al*: At-TAX: a whole genome tiling array resource for developmental expression analysis and transcript identification in Arabidopsis thaliana. *Genome Biol* 2008, **9**:R112.
- Naouar N, Vandepoele K, Lammens T, Casneuf T, Zeller G, van Hummelen P, Weigel D, Ratsch G, Inze D, Kuiper M, *et al*: Quantitative RNA expression analysis with Affymetrix Tiling 1.0R arrays identifies new E2F target genes. *Plant J* 2008, **57**:184–194.
- Zhang X, Shiu S, Cal A, Borevitz JO: Global analysis of genetic, epigenetic and transcriptional polymorphisms in Arabidopsis thaliana using whole genome tiling arrays. *PLoS Genet* 2008, **4**:e1000032.
- Hazen SP, Naef F, Quisel T, Gendron JM, Chen H, Ecker JR, Borevitz JO, Kay SA: Exploring the transcriptional landscape of plant circadian rhythms using genome tiling arrays. *Genome Biol* 2009, **10**:R17.
- He G, Elling AA, Deng XW: The epigenome and plant development. *Annu Rev Plant Biol* 2011, **62**:411–435.
- Nagano AJ, Fukazawa M, Hayashi M, Ikeuchi M, Tsukaya H, Nishimura M, Hara-Nishimura I: AtMap1: a DNA microarray for genomic deletion mapping in Arabidopsis thaliana. *Plant J* 2008, **56**:1058–1065.
- Jones-Rhoades MW, Borevitz JO, Preuss D: Genome-wide expression profiling of the Arabidopsis female gametophyte identifies families of small, secreted proteins. *PLoS Genet* 2007, **3**:1848–1861.
- Richardson CR, Luo QJ, Gontcharova V, Jiang YW, Samanta M, Youn E, Rock CD: Analysis of antisense expression by whole genome tiling microarrays and siRNAs suggests mis-annotation of Arabidopsis orphan protein-coding genes. *PLoS One* 2010, **5**:e10710.
- Oh S, Park S, van Nocker S: Genic and global functions for Paf1C in chromatin modification and gene expression in Arabidopsis. *PLoS Genet* 2008, **4**:e1000077.
- Kaufmann K, Muino JM, Jauregui R, Airoldi CA, Smaczniak C, Krajewski P, Angenent GC: Target genes of the MADS transcription factor SEPALLATA3: integration of developmental and hormonal pathways in the Arabidopsis flower. *PLoS Biol* 2009, **7**:e1000090.
- Matsui A, Ishida J, Morosawa T, Mochizuki Y, Kaminuma E, Endo TA, Okamoto M, Nambara E, Nakajima M, Kawashima M, *et al*: Arabidopsis transcriptome analysis under drought, cold, high-salinity and ABA treatment conditions using a tiling array. *Plant Cell Physiol* 2008, **49**:1135–1149.
- Rehrauer H, Aquino C, Gruissem W, Henz SR, Hilson P, Laubinger S, Naouar N, Patrignani A, Rombauts S, Shu H, *et al*: AGRONOMICS1: A new resource for Arabidopsis transcriptome profiling. *Plant Physiol* 2010, **152**:487–499.
- Massonnet C, Vile D, Fabre J, Hannah MA, Caldana C, Lisee J, Beemster GT, Meyer RC, Messerli G, Gronlund JT, *et al*: Probing the reproducibility of leaf growth and molecular phenotypes: a comparison of three Arabidopsis accessions cultivated in ten laboratories. *Plant Physiol* 2010, **152**:2142–2157.
- Weinhofer I, Hehenberger E, Roszak P, Hennig L, Köhler C: H3K27me3 profiling of the endosperm implies exclusion of Polycomb group protein targeting by DNA methylation. *PLoS Genet* 2010, **6**:e1001152.
- Bischof S, Baerenfaller K, Wildhaber T, Troesch R, Vidi PA, Roschitzki B, Hirsch-Hoffmann M, Hennig L, Kessler F, Gruissem W, *et al*: Plastid proteome assembly without Toc159: Photosynthetic protein import and accumulation of n-acetylated plastid precursor proteins. *Plant Cell* 2011, **23**:3911–3928.
- Slomovic S, Portnoy V, Schuster G: RNA Polyadenylation in prokaryotes and organelles; Different tails tell different tales. *Crit Rev Plant Sci* 2006, **25**:65–77.
- Nelson DC, Wohlbach DJ, Rodesch MJ, Stolc V, Sussman MR, Samanta MP: Identification of an in vitro transcription-based artifact affecting oligonucleotide microarrays. *FEBS Lett* 2007, **581**:3363–3370.
- Xiang CC, Kozhich OA, Chen M, Inman JM, Phan QN, Chen Y, Brownstein MJ: Amine-modified random primers to label probes for DNA microarrays. *Nat Biotechnol* 2002, **20**:738–742.
- Pio R, Blanco D, Pajares MJ, Aibar E, Durany O, Ezponda T, Agorreta J, Gomez-Roman J, Anton MA, Rubio A, *et al*: Development of a novel splice array platform and its application in the identification of alternative splice variants in lung cancer. *BMC Genomics* 2010, **11**:352.
- Fasold M, Stadler PF, Binder H: G-stack modulated probe intensities on expression arrays - sequence corrections and signal calibration. *BMC Bioinformatics* 2010, **11**:207.
- Cope L, Hartman SM, Gohlmann HW, Tiesman JP, Irizarry RA: Analysis of Affymetrix GeneChip data using amplified RNA. *Biotechniques* 2006, **40**:165–170.
- Stangegaard M, Dufva IH, Dufva M: Reverse transcription using random pentadecamer primers increases yield and quality of resulting cDNA. *Biotechniques* 2006, **40**:649–657.
- Coram TE, Settles ML, Chen X: Large-scale analysis of antisense transcription in wheat using the Affymetrix GeneChip Wheat Genome Array. *BMC Genomics* 2009, **10**:253.

36. Huber W, Toedling J, Steinmetz LM: **Transcript mapping with high-density oligonucleotide tiling arrays.** *Bioinformatics* 2006, **22**:1963–1970.
37. Kim BS, Rha SY, Cho GB, Chung HC: **Spearman's footrule as a measure of cDNA microarray reproducibility.** *Genomics* 2004, **84**:441–448.
38. R Development Core Team: *R: A language and environment for statistical computing.* Vienna, Austria: R Foundation for Statistical Computing; 2010.
39. Nicol JW, Helt GA, Blanchard SG Jr, Raja A, Loraine AE: **The Integrated Genome Browser: free software for distribution and exploration of genome-scale datasets.** *Bioinformatics* 2009, **25**:2730–2731.
40. Irizarry RA, Bolstad BM, Collin F, Cope LM, Hobbs B, Speed TP: **Summaries of Affymetrix genechip probe level data.** *Nucleic Acids Res* 2003, **31**:e15.
41. Bengtsson H, Simpson K, Bullard J, Hansen K: *aroma.affymetrix: A generic framework in R for analyzing small to very large Affymetrix data sets in bounded memory.* Berkeley: Tech Report #745, Department of Statistics, University of California; 2008.
42. Smyth GK: **Linear models and empirical Bayes methods for assessing differential expression in microarray experiments.** *Stat Appl Genet Mol Biol* 2004, **3**:1–26.

doi:10.1186/1746-4811-8-18

Cite this article as: Müller *et al.*: Evaluation of alternative RNA labeling protocols for transcript profiling with Arabidopsis AGRONOMICS1 tiling arrays. *Plant Methods* 2012 **8**:18.

Submit your next manuscript to BioMed Central and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit

